

Integrasi Principal Component Analysis dan DBSCAN untuk Klasterisasi Tingkat Keparahan Stunting pada Balita

An Integrated Principal Component Analysis and DBSCAN Approach for Clustering Stunting Severity Levels among Toddlers

Azzah Aulia Syarif^{a,1}, Rizki Yusliana Bakti^{a,2*}, Muhammad Faisal^{a,3}, Nini Apriani Rumata^{b,4}, Titik Khawa Abd Rahman^{c,5}, Andi Makbul Syamsuri^{d,6}, dan M. Aguslim^{d,7}

^a Department of Informatics, Universitas Muhammadiyah Makassar, Makassar, Indonesia
^b Regional and Urban Planning, Universitas Muhammadiyah Makassar, Makassar, Indonesia
^c School Science and Technology, Asia E University, Selangor, Malaysia
^d Water Resources Engineering, Universitas Muhammadiyah Makassar, Makassar, Indonesia
¹ azzahauliasyarif04@gmail.com; ² rizkiyusliana@unismuh.ac.id; ³ muhfaisal@unismuh.ac.id;
⁴ nini.rumata@unismuh.ac.id; ⁵ titik.khawa@aeu.edu.my; ⁶ amakbulsyamsuri@unismuh.ac.id;
⁷ m.aguslim@unismuh.ac.id
*corresponding author

Informasi Artikel	ABSTRAK
Diserahkan : 12 Mei 2026 Diterima : 15 Agustus 2026 Direvisi : 22 Agustus 2026 Diterbitkan : 26 Agustus 2026 Kata Kunci: Stunting Status Gizi Balita Klasterisasi Tanpa Pengawasan Principal Component Analysis (PCA) DBSCAN	Stunting masih menjadi masalah kesehatan masyarakat yang serius. Variasi karakteristik antropometri pada balita stunting membuat satu kategori klasterisasi tidak cukup merepresentasikan kondisi pertumbuhan secara menyeluruh. Penelitian ini mengusulkan integrasi <i>Principal Component Analysis</i> (PCA) dan DBSCAN untuk mengidentifikasi subkelompok keparahan stunting pada balita di Puskesmas Turikale (Agustus 2025; n = 406). Setelah pembersihan dan standarisasi usia, PCA diterapkan pada HAZ, WAZ, WHZ, dan usia saat pengukuran, menghasilkan tiga komponen utama (PC1–PC3) yang menjelaskan 99.35% variasi data. DBSCAN pada ruang PCA ($\epsilon = 0.60$; MinPts = 5; Euclidean) menghasilkan tiga klaster dengan 12.56% noise yang direalokasi menggunakan <i>nearest-centroid</i> agar seluruh sampel berlabel. Hasil menunjukkan tiga subkelompok: stunting sedang dengan underweight pada usia lebih tua (n = 357), stunting sedang disertai risiko gizi lebih pada usia lebih muda (n = 24), dan stunting berat pada usia dini (n = 25). Dibandingkan K-medoids dan Fuzzy C-Means, PCA–DBSCAN memberikan validasi internal terbaik (Silhouette = 0.277; DBI = 1.062; Dunn = 0.0765). Pendekatan ini menghasilkan profil antropometri yang dapat digunakan sebagai alat bantu penentuan prioritas pemantauan dan penilaian gizi di layanan kesehatan primer, tanpa menggantikan klasifikasi WHO.
Keywords: Stunting Unsupervised clustering Principal Component Analysis (PCA) DBSCAN Anthropometric Data	ABSTRACT <i>Stunting remains a serious public health issue. The variation in anthropometric characteristics among stunted toddlers means that a single clustering category is insufficient to fully represent their growth status. This study proposes a hybrid approach combining integrated Principal Component Analysis (PCA) and DBSCAN to identify subgroups of stunting severity among toddlers at the Turikale Community Health Center (August 2025; n = 406). After data cleaning and z-score standardization of age at measurement, PCA was applied to HAZ, WAZ, WHZ, and age at measurement, yielding three principal components (PC1–PC3) that explained 99.35% of the data variation. DBSCAN in the PCA space ($\epsilon = 0.60$; MinPts = 5; Euclidean) yielded three clusters with 12.56% noise, which was reallocated using nearest-centroid to ensure all samples were labeled. The results revealed three subgroups: moderate stunting with underweight at older ages (n = 357), moderate stunting with possible risk of overweight at younge ages (n = 24), and severe stunting at early ages (n = 25). Compared to K-medoids and Fuzzy C-Means, PCA–DBSCAN provided the best internal validation (Silhouette = 0.277; DBI = 1.062; Dunn = 0.0765). This approach provides interpretable anthropometric profiles that may support the prioritization of growth monitoring and nutritional assessment in primary health care, without replacing WHO classification</i>

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



I. Pendahuluan

Menurut World Health Organization (WHO), stunting pada anak di bawah lima tahun didefinisikan berdasarkan indeks *height-for-age*, yaitu nilai *height-for-age z-score* (HAZ) < -2 SD dari median *WHO Child Growth Standards* [1], [2]. Stunting merupakan salah satu masalah kesehatan masyarakat yang signifikan karena berdampak luas terhadap perkembangan fisik, kognitif, serta kualitas sumber daya manusia suatu negara [3]. Secara biologis, kondisi ini mencerminkan kekurangan gizi kronis serta paparan infeksi berulang pada awal kehidupan, yang menyebabkan tinggi badan anak berada di bawah standar usianya [4]. Meskipun terjadi kemajuan global dalam satu dekade terakhir, sekitar 148 juta ($\pm 22\%$) anak di bawah lima tahun masih mengalami stunting pada tahun 2022, dengan sebagian besar kasus terkonsentrasi di Afrika dan Asia [5].

Indonesia termasuk negara dengan prevalensi tinggi, sehingga pemerintah menetapkan Peraturan Presiden Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting sebagai upaya strategis nasional [6], [7]. Di Kabupaten Maros, prevalensi stunting menurun dari 34.7% pada tahun 2023 menjadi 22.4% pada Juli 2025. Namun, berdasarkan data e-PPGBM, Kecamatan Turikale mencatat 529 kasus stunting dan menjadi kecamatan dengan jumlah kasus stunting tertinggi kedua di Kabupaten Maros [8]. Penanganan stunting di Kabupaten Maros dilaksanakan melalui kolaborasi lintas sektor yang meliputi pendataan melalui e-PPGBM, pemberian makanan tambahan, standarisasi alat antropometri, pemantauan pertumbuhan bulanan melalui Posyandu, percepatan rujukan bagi balita dengan penyakit penyerta, penyediaan air minum dan perbaikan sanitasi, pendampingan keluarga, suplementasi gizi remaja, serta pencegahan pernikahan dini [9]. Kondisi tersebut menunjukkan perlunya pemetaan balita stunting untuk mendukung pemantauan dan tindak lanjut gizi yang lebih terarah.

Penetapan stunting berdasarkan TB/U atau HAZ belum sepenuhnya merepresentasikan kondisi antropometri anak. Penelitian di Indonesia menunjukkan bahwa stunting dapat terjadi sebagai kegagalan antropometri tunggal maupun bersamaan dengan masalah gizi lainnya. Permatasari dan Chadirin [10] menemukan kategori stunting saja, stunting disertai *underweight*, serta kombinasi stunting, *underweight*, dan *wasting*. Jokhu dan Syauqy [11] melaporkan kejadian stunting dan *wasting* secara bersamaan pada anak yang sama, sedangkan Paramashanti et al. [12] mengidentifikasi kejadian stunting yang disertai kelebihan berat badan. Temuan tersebut menunjukkan bahwa balita yang sama-sama terklasifikasi stunting dapat memiliki profil antropometri dan kebutuhan tindak lanjut gizi yang berbeda. Namun, subkelompok balita stunting berdasarkan pola gabungan HAZ, WAZ, WHZ, dan usia belum dipetakan di Puskesmas Turikale. Pemetaan ini diharapkan dapat mendukung identifikasi balita dengan defisit antropometri ganda yang memerlukan penilaian gizi dan pemantauan lebih intensif, tanpa menggantikan klasifikasi WHO maupun penilaian klinis tenaga kesehatan.

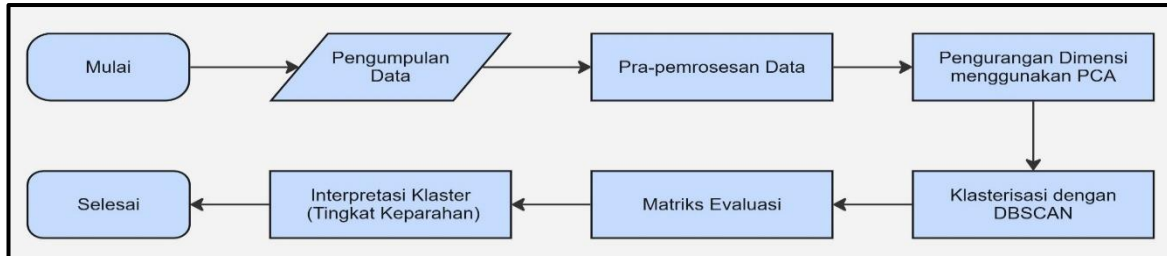
Berbagai penelitian menunjukkan bahwa stunting dipengaruhi oleh faktor multidimensi, termasuk faktor sosial ekonomi, kondisi lingkungan fisik dan iklim, serta karakteristik rumah tangga seperti riwayat infeksi berulang, pendidikan ibu, ketahanan pangan, dan praktik pemberian makan [13], [14]. Namun, regresi logistik dan pendekatan analisis konvensional berbasis model linier memiliki keterbatasan dalam memodelkan hubungan yang kompleks dan nonlinier [15]. Selain itu, metode klusterisasi seperti *hierarchical clustering* dan K-Means kurang optimal untuk data dengan distribusi tidak merata, mengandung *noise*, serta memiliki batas antar kelompok yang tidak jelas [16]. Oleh karena itu, masih terdapat keterbatasan metode analisis untuk mengidentifikasi variasi kondisi stunting secara komprehensif, terutama pada data antropometri multivariat yang dapat mengandung *outlier* dan memiliki distribusi kepadatan yang tidak merata. Untuk mengatasi keterbatasan tersebut, pendekatan *unsupervised learning* menjadi alternatif yang efektif. Beberapa studi telah mengintegrasikan *Principal Component Analysis* (PCA) untuk reduksi dimensi dengan metode klusterisasi [17]. Di sisi lain, *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) merupakan algoritma berbasis kepadatan yang tidak memerlukan penentuan jumlah kluster di awal serta mampu mengidentifikasi data sebagai *noise* atau *outlier* [18]. Namun, pemanfaatan DBSCAN dalam analisis profil stunting, khususnya melalui integrasi dengan PCA pada data layanan kesehatan primer, masih relatif terbatas. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan metode klusterisasi berbasis partisi atau hierarki, penelitian ini mengusulkan integrasi PCA–DBSCAN untuk menganalisis data antropometri multivariat serta mengungkap pola laten dan heterogenitas profil stunting pada populasi balita.

Berpedoman pada kategori keparahan WHO yang membedakan stunting sedang dan berat [2], penelitian ini menghipotesiskan terbentuknya sedikitnya dua subkelompok yang dapat diinterpretasikan secara klinis. Namun, jumlah kluster akhir tidak ditetapkan sebelumnya, melainkan ditentukan oleh struktur kepadatan data menggunakan DBSCAN [19]. Penelitian ini bertujuan untuk: (1) menerapkan PCA dalam mereduksi dimensi dan mengurangi korelasi antarvariabel pada data balita stunting, (2) menerapkan algoritma DBSCAN untuk menghasilkan kluster profil pada populasi balita yang telah terklasifikasi stunting berdasarkan kemiripan pola indikator antropometri, dan (3) mengevaluasi kualitas kluster PCA–DBSCAN dan membandingkannya dengan K-Medoids dan Fuzzy C-Means menggunakan *Silhouette Score*, *Davies-Bouldin Index*, dan *Dunn Index*, serta menginterpretasikan karakteristik klinis masing-masing kluster.

II. Metode

A. Kerangka Penelitian

Kerangka penelitian ini disusun untuk menyajikan alur metodologi secara sistematis, sebagaimana ditunjukkan pada Gambar 1. Kerangka ini memastikan bahwa setiap tahapan mulai dari pengumpulan data hingga interpretasi hasil kluster dilaksanakan secara terstruktur sebagai satu rangkaian proses analisis yang terintegrasi.



Gambar 1. Kerangka Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan pengumpulan data balita stunting usia 0–59 bulan di Puskesmas Turikale. Tahapan selanjutnya adalah pra-pemrosesan data yang meliputi pembersihan data, penanganan data hilang, pengodean variabel kategorik, evaluasi *outlier* berdasarkan kriteria *biologically implausible values* (BIV), serta standardisasi data.

Data selanjutnya direduksi menggunakan *Principal Component Analysis* (PCA). Hasil reduksi dimensi kemudian digunakan sebagai input dalam proses klusterisasi menggunakan metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Tahap akhir meliputi *post-processing* dan evaluasi kualitas kluster menggunakan *Silhouette Score*, *Davies–Bouldin Index* (DBI), dan *Dunn Index*.

B. Wilayah Studi dan Sumber Data

Penelitian ini dilakukan di wilayah kerja Puskesmas Turikale, Kabupaten Maros, Provinsi Sulawesi Selatan. Puskesmas ini merupakan fasilitas kesehatan tingkat pertama yang membawahi beberapa posyandu dan kelurahan [20].

Data yang digunakan merupakan data sekunder yang diperoleh dari rekam medis balita stunting usia 0–59 bulan yang tercatat dalam register pemantauan pertumbuhan dan status gizi pada periode Agustus 2025, dengan jumlah sampel sebanyak 406 balita.

Data yang dianalisis mencakup informasi antropometri dan indikator status gizi. Permohonan pelaksanaan penelitian disampaikan melalui Surat LP3M Universitas Muhammadiyah Makassar Nomor 831/LP3M/05/C.4-VIII/X/1447/2025 tanggal 22 Oktober 2025 yang ditujukan kepada Dinas Kesehatan Kabupaten Maros. Berdasarkan surat tersebut, akses terhadap data difasilitasi oleh Puskesmas Turikale. Dataset dianalisis dalam bentuk terdeidentifikasi tanpa menyertakan nama, alamat, maupun pengenalan langsung lainnya. Penelitian tidak melibatkan kontak langsung atau intervensi terhadap balita.

C. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum analisis [21]. Proses diawali dengan penghapusan atribut yang tidak relevan serta berpotensi melanggar kerahasiaan data. Selanjutnya dilakukan pemeriksaan nilai hilang pada seluruh variabel. Hasil pemeriksaan menunjukkan tidak terdapat nilai hilang, sehingga proses imputasi tidak diperlukan.

Interpretasi status gizi mengacu pada standar *World Health Organization* (WHO) berbasis *z-score*, yaitu kategori sedang pada rentang $-3 \leq z\text{-score} < -2$ dan kategori berat pada $z\text{-score} < -3$ untuk HAZ (stunting), WAZ (underweight), dan WHZ (wasting) [22]. Evaluasi *outlier* dilakukan menggunakan kriteria *biologically implausible values* (BIV), yaitu HAZ < -6 atau $> +6$, WAZ < -6 atau $> +5$, dan WHZ < -5 atau $> +5$ [23]. Karena variabel HAZ, WAZ, dan WHZ telah berada pada skala *z-score*, proses standardisasi hanya diterapkan pada variabel usia menggunakan Persamaan (1).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

z : Nilai hasil standardisasi

x : Usia saat pengukuran

μ : Rata-rata usia

σ : Simpangan baku usia

D. Pengurangan Dimensi dengan Principal Component Analysis (PCA)

Pengurangan dimensi dilakukan menggunakan *Principal Component Analysis* (PCA) untuk mengubah variabel yang saling berkorelasi menjadi sejumlah komponen utama yang tidak saling berkorelasi, sehingga

dapat mengurangi redundansi data tanpa menghilangkan informasi penting [24]. PCA diterapkan pada empat fitur berdasarkan WHO, yaitu HAZ (ZS TB/U), WAZ (ZS BB/U), WHZ (ZS BB/TB) [25], serta usia saat pengukuran yang telah distandarisasi pada tahap sebelumnya.

Jumlah komponen utama yang dipertahankan ditentukan berdasarkan *explained variance ratio* dan *cumulative explained variance* [26], dengan kriteria varians kumulatif $\geq 95\%$. Dua komponen pertama hanya menjelaskan 86,51% varians sehingga belum memenuhi ambang tersebut. Penambahan PC3 meningkatkan varians kumulatif menjadi 99,35%. Oleh karena itu, tiga komponen dipilih sebagai jumlah minimum yang memenuhi kriteria retensi informasi.

E. Klasterisasi dengan DBSCAN

Klasterisasi dilakukan menggunakan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) karena mampu membentuk kluster berbasis kepadatan dengan bentuk *arbitrer*, mengidentifikasi *noise*, dan tidak memerlukan penetapan jumlah kluster di awal [27]. Pada penelitian ini, DBSCAN diterapkan pada skor hasil PCA (PC1–PC3) menggunakan metrik jarak *Euclidean*.

Penentuan parameter epsilon (ϵ) dilakukan melalui uji sensitivitas menggunakan *k-distance plot* dengan pendekatan metode siku (*elbow method*) pada nilai $k = 5$, sedangkan parameter MinPts ditentukan dengan membandingkan beberapa nilai kandidat berdasarkan metrik validasi internal.

Berdasarkan parameter yang digunakan, titik data dengan jumlah tetangga $\geq \text{MinPts}$ dalam radius ϵ dikategorikan sebagai *core point*. Titik yang tidak memenuhi kriteria tersebut tetapi masih dapat dijangkau (*density-reachable*) dari *core point*, dikategorikan sebagai *border point*. Sementara itu, titik lainnya dianggap sebagai *noise* [28]. Hasil proses klasterisasi berupa label kluster, dengan label -1 yang menunjukkan data yang tidak termasuk dalam kluster mana pun.

Untuk memastikan seluruh sampel tercakup dalam pemetaan profil keparahan stunting, dilakukan tahap *post-processing* berupa konsolidasi label kluster setelah DBSCAN. Tiga kluster *non-noise* dengan jumlah anggota terbesar dipertahankan dan identitasnya dipetakan ulang menjadi label $\{0, 1, 2\}$. Data berlabel -1 tidak dipertahankan sebagai kategori tersendiri karena label tersebut hanya menunjukkan bahwa suatu titik tidak memenuhi kriteria kepadatan DBSCAN, bukan menunjukkan bahwa seluruh titik *noise* membentuk satu kelompok dengan karakteristik yang sama. Mempertahankan seluruh *noise* sebagai satu kategori berpotensi menggabungkan titik-titik yang berada pada lokasi berbeda dalam ruang PCA. Oleh karena itu, untuk memenuhi tujuan pemetaan seluruh sampel, pendekatan *nearest centroid* dipilih sebagai aturan objektif dan deterministik dengan mengalokasikan setiap *noise* ke kluster yang centroidnya memiliki jarak Euclidean paling kecil. *Centroid* setiap kluster dihitung dari rata-rata koordinat PC1–PC3 anggotanya sebelum realokasi. Prosedur serupa diterapkan dalam ECR-DBSCAN dengan mengalokasikan *noise* ke kluster yang memiliki jarak minimum terhadap centroid [29]. Realokasi dilakukan setelah pembentukan kluster awal sehingga tidak memengaruhi proses pembentukan kluster awal oleh DBSCAN.

F. Metrik Evaluasi

Kualitas pemisahan kluster PCA–DBSCAN dievaluasi dan dibandingkan dengan hasil klasterisasi K-Medoids dan Fuzzy C-Means menggunakan tiga metrik validasi internal, yaitu *Silhouette Score*, *Davies–Bouldin Index* (DBI), dan *Dunn Index* [30].

1. Silhouette Score

Silhouette Score mengukur kesesuaian setiap sampel terhadap kluster tempatnya berada dengan membandingkannya terhadap kluster terdekat [31]. Nilai *Silhouette Score* berada pada rentang $[-1, 1]$ [32]. Nilai yang semakin mendekati 1 menunjukkan pemisahan kluster yang semakin baik. Untuk setiap sampel i , koefisien *Silhouette Score* didefinisikan pada Persamaan (2):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

$a(i)$: rata-rata jarak sampel i terhadap seluruh anggota dalam kluster yang sama

$b(i)$: rata-rata jarak minimum sampel i terhadap kluster lain terdekat

2. Davies–Bouldin Index

Davies–Bouldin Index mengukur kualitas kluster berdasarkan rasio antara tingkat penyebaran data dalam kluster dan jarak antarkluster [33]. Semakin rendah nilai DBI, semakin baik kualitas kluster yang dihasilkan. DBI didefinisikan pada Persamaan (3):

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} R_{ij} \quad (3)$$

R_{ij} : tingkat kemiripan antara kluster ke- i dan kluster ke- j , yang dihitung dalam Persamaan (4):

$$R_{ij} = \left(\frac{s_i + s_j}{d_{ij}} \right) \quad (4)$$

s_i dan s_j : tingkat penyebaran data pada kluster ke- i dan kluster ke- j

3. Dunn Index

Dunn Index didefinisikan sebagai rasio antara jarak minimum antarkluster dan tingkat penyebaran maksimum dalam kluster [34]. Nilai *Dunn Indeks* yang lebih tinggi menunjukkan kluster yang semakin terpisah dan semakin kompak [35]. *Dunn Index* didefinisikan pada Persamaan (5):

$$D(k) = \frac{\min_{1 \leq r < s \leq k} I(C_r, C_s)}{\max_{1 \leq r < s \leq k} \Delta_q} \quad (5)$$

Δ_q : diameter kluster ke- q

$I(C_r, C_s)$: jarak antar-kluster antara kluster C_r dan C_s

G. Analisis Statistik Antarkluster

Perbedaan karakteristik HAZ, WAZ, WHZ, dan usia antara ketiga kluster dianalisis menggunakan uji Kruskal–Wallis. Variabel yang menunjukkan perbedaan signifikan ($p < 0.05$) selanjutnya dianalisis menggunakan uji *Post-Hoc Dunn-Bonferroni* dengan koreksi Bonferroni untuk mengidentifikasi pasangan kluster yang berbeda secara signifikan [36]. Besarnya perbedaan antarkluster dilaporkan menggunakan eta-squared (η^2) sebagai ukuran efek uji Kruskal–Wallis [37].

III. Hasil dan Pembahasan

A. Gambaran Data

Untuk memberikan gambaran karakteristik data, Tabel 1 menyajikan statistik deskriptif variabel berupa nilai rata-rata (mean), standar deviasi, minimum, dan maksimum untuk variabel HAZ, WAZ, WHZ, serta usia saat pengukuran. Ketiga indikator antropometri telah dinyatakan dalam satuan *z-score* WHO, sedangkan usia saat pengukuran telah distandardisasi ke skala standar sebelum proses PCA.

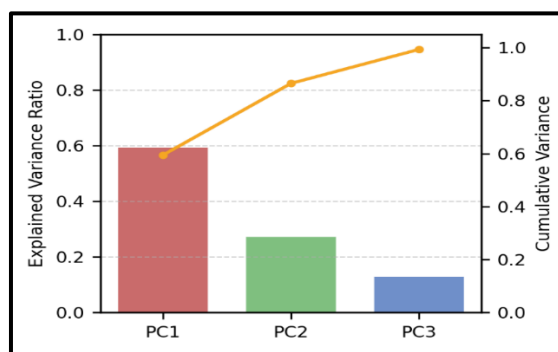
Tabel 1. Statistik Deskriptif Variabel

Indikator	Mean	SD	Min	Max
HAZ	-2.74	0.62	-5.63	-2.01
WAZ	-1.97	0.88	-4.86	1.17
WHZ	-0.58	1.19	-3.49	4.64
Usia Saat Pengukuran	0.00	1.00	-1.97	2.07

B. Hasil Principal Component Analysis (PCA)

Tiga komponen utama (PC1–PC3) mampu merepresentasikan hampir seluruh variasi data. Nilai *explained variance ratio* masing-masing komponen adalah 59.42% untuk PC1, 27.09% untuk PC2, dan 12.85% untuk PC3, sehingga total variansi kumulatif yang dijelaskan mencapai 99.35%. Oleh karena itu, ketiga komponen utama tersebut digunakan sebagai representasi data pada tahap klusterisasi.

Proporsi variansi yang dijelaskan oleh setiap komponen utama ditunjukkan pada Gambar 2, yang menampilkan *explained variance ratio* dalam bentuk diagram batang dan variansi kumulatif dalam bentuk kurva.



Gambar 2. Scree Plot *Explained Variance Ratio* dan Variansi Kumulatif (PC1–PC3)

Selanjutnya, interpretasi setiap komponen dilakukan menggunakan nilai *Scaled Loading* yang disajikan pada Tabel 2. Komponen utama pertama (PC1) terutama dibentuk oleh WHZ (1,17) dan WAZ (0,81), dengan nilai terbesar pada WHZ, sedangkan usia saat pengukuran memiliki loading negatif (−0,31). Hasil tersebut menunjukkan bahwa PC1 merepresentasikan variasi status berat badan terhadap tinggi badan. Komponen utama kedua (PC2) didominasi oleh usia saat pengukuran (0,94), sehingga merepresentasikan dimensi usia. Sementara itu, komponen utama ketiga (PC3) terutama dibentuk oleh HAZ (0,59), sehingga merepresentasikan variasi pertumbuhan linier yang berkaitan langsung dengan stunting.

Nilai WHZ sebesar 1,17 pada PC1 dapat melebihi 1 karena nilai yang disajikan merupakan *eigenvalue-scaled loading*, yang dihitung dengan mengalikan *eigenvector* dengan akar *eigenvalue*. Oleh karena itu, nilainya tidak dibatasi pada rentang -1 hingga $+1$.

Temuan ini menunjukkan bahwa PCA mampu merangkum variasi utama yang berkaitan dengan status berat badan, usia, dan pertumbuhan linier ke dalam komponen yang berbeda sebelum proses klusterisasi dilakukan.

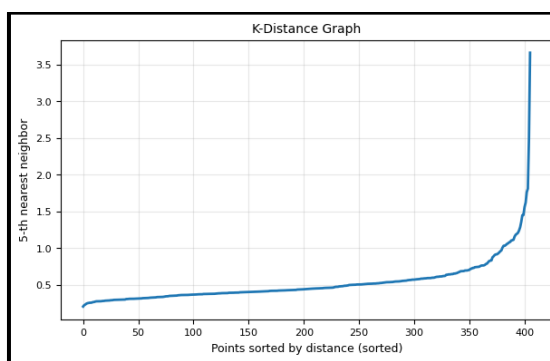
Tabel 2. Nilai *Scaled Loading* PCA

Indikator	PC1	PC2	PC3
HAZ	-0.01	0.16	0.59
WAZ	0.81	0.19	0.24
WHZ	1.17	0.12	-0.19
Usia Saat Pengukuran	-0.31	0.94	-0.12

C. Hasil Klusterisasi dengan DBSCAN

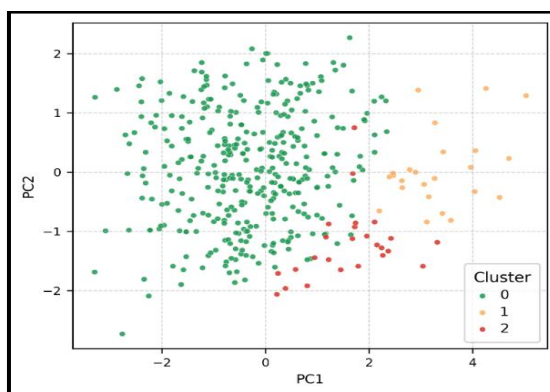
DBSCAN diterapkan pada skor PCA (PC1–PC3) menggunakan metrik jarak *Euclidean*. Berdasarkan Gambar 3, *k-distance plot* ($k = 5$) menunjukkan titik siku pada nilai $\epsilon \approx 0.60$. Oleh karena itu, nilai $\epsilon = 0.60$ dipilih karena mampu membentuk pola kluster data secara optimal sekaligus mengidentifikasi titik data yang menyimpang dari pola pengelompokan data sebagai *noise* atau *outlier*.

Parameter MinPts ditentukan melalui uji sensitivitas menggunakan *Silhouette Score*, *Davies–Bouldin Index* (DBI), dan *Dunn Index* [38], [39]. Hasil evaluasi menunjukkan bahwa nilai MinPts = 5 memberikan konfigurasi kluster terbaik karena menghasilkan pemisahan kluster dan stabilitas yang lebih baik dibandingkan nilai parameter lainnya.



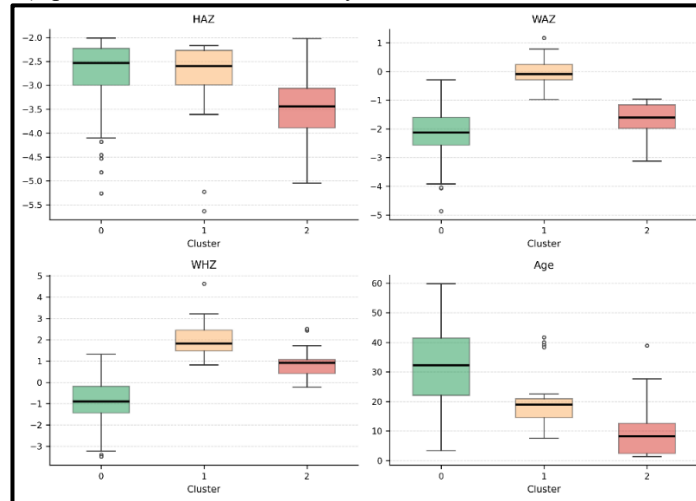
Gambar 3. Penentuan Nilai Epsilon (ϵ) pada DBSCAN

Dengan konfigurasi $\epsilon = 0.60$ dan MinPts = 5, DBSCAN menghasilkan tiga kluster *non-noise* serta mengidentifikasi 51 sampel (12.56%) sebagai *noise* (label -1). Distribusi kluster awal terdiri atas kluster 0 sebanyak 336 sampel (82.76%), kluster 1 sebanyak 11 sampel (2.71%), dan kluster 2 sebanyak 8 sampel (1.97%). Selanjutnya, sampel dengan label -1 direlokasikan menggunakan pendekatan *nearest centroid* pada ruang PCA sehingga seluruh sampel memperoleh label kluster. Setelah proses realokasi, distribusi kluster akhir menjadi kluster 0 sebanyak 357 sampel (87.93%), kluster 1 sebanyak 24 sampel (5.91%), dan kluster 2 sebanyak 25 sampel (6.16%). Proyeksi sebaran kluster pada bidang PC1–PC2 ditunjukkan pada Gambar 4.



Gambar 4. Proyeksi Sebaran Kluster DBSCAN pada Bidang PC1–PC2

Berdasarkan nilai median tiap indikator pada Gambar 5, kluster 0 menunjukkan karakteristik stunting sedang ($HAZ = -2.53$) yang disertai kondisi *underweight* ($WAZ = -2.12$), dengan nilai WHZ terendah di antara ketiga kluster (-0.90) serta median usia tertua, yaitu 32.27 bulan. Kluster 1 juga merepresentasikan stunting sedang ($HAZ = -2.60$), tetapi dengan status berat badan yang mendekati normal ($WAZ = -0.08$) dan nilai WHZ yang relatif tinggi (1.83), pada median usia 18.92 bulan. Adapun kluster 2 mengindikasikan stunting berat ($HAZ = -3.44$) dengan status berat badan menurut usia sebesar -1.61 dan nilai WHZ yang masih berada pada rentang normal (0.91), pada median usia termuda, yaitu 8.20 bulan.



Gambar 5. Boxplot HAZ, WAZ, WHZ, dan Usia Saat Pengukuran berdasarkan Kluster DBSCAN

Dengan demikian, ketiga kluster dapat diklasifikasikan sebagai berikut: Kluster 0 merupakan kelompok balita dengan stunting sedang disertai *underweight* pada usia relatif lebih tua. Kluster 1 merupakan kelompok balita dengan stunting sedang disertai risiko gizi lebih yang dominan pada usia relatif lebih muda; dan kluster 2 merupakan kelompok balita dengan stunting berat pada usia dini.

D. Analisis Statistik Antarkluster

Uji Kruskal–Wallis dilakukan untuk memverifikasi perbedaan karakteristik HAZ, WAZ, WHZ, dan usia saat pengukuran antara ketiga kluster akhir. Hasil pengujian pada Tabel 3 menunjukkan bahwa seluruh variabel berbeda secara signifikan antarkluster ($p < 0.001$).

Tabel 3. Hasil Uji Kruskal–Wallis

Indikator	H	df	p	η^2	Efek
HAZ	25.067	2	<0.001	0.057	Kecil
WAZ	73.860	2	<0.001	0.178	Besar
WHZ	114.386	2	<0.001	0.279	Besar
Usia Saat Pengukuran	63.369	2	<0.001	0.152	Besar

Hasil ini menunjukkan bahwa perbedaan karakteristik antarkluster tidak hanya terlihat secara deskriptif pada boxplot, tetapi juga teridentifikasi secara statistik. Uji post-hoc Dunn dengan koreksi Bonferroni selanjutnya dilakukan untuk mengetahui pasangan kluster yang berbeda secara signifikan, sebagaimana disajikan pada Tabel 4.

Tabel 4. Ringkasan Hasil Uji Post-Hoc Dunn–Bonferroni

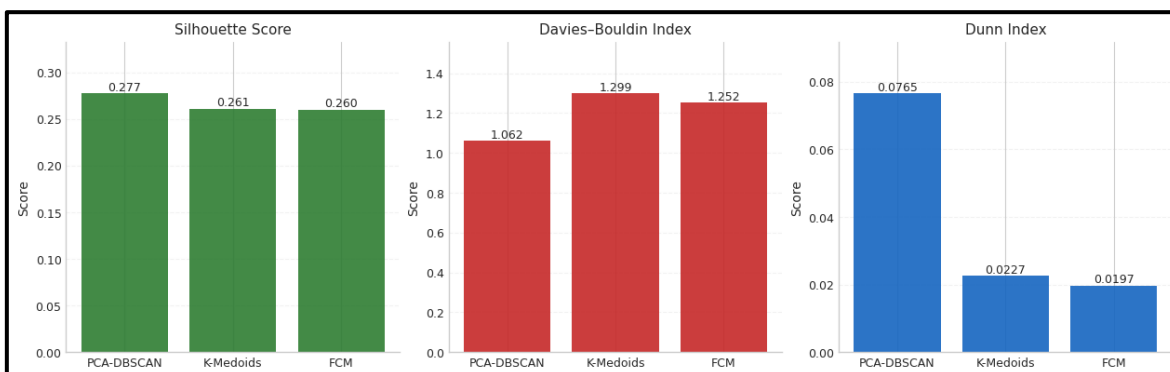
Indikator	Kluster 0 vs 1	Kluster 0 vs 2	Kluster 1 vs 2
HAZ	1.000 (Tidak signifikan)	<0.001 (Signifikan)	0.011 (Signifikan)
WAZ	<0.001 (Signifikan)	0.020 (Signifikan)	<0.001 (Signifikan)
WHZ	<0.001 (Signifikan)	<0.001 (Signifikan)	0.763 (Tidak signifikan)
Usia Saat Pengukuran	<0.001 (Signifikan)	<0.001 (Signifikan)	0.067 (Tidak signifikan)

Catatan: Nilai yang disajikan merupakan nilai p yang telah dikoreksi menggunakan metode Bonferroni. Perbedaan dinyatakan signifikan apabila $p < 0.05$.

Uji *Post - Hoc Dunn–Bonferroni* menunjukkan bahwa HAZ kluster 2 berbeda signifikan dari kluster 0 dan 1, sedangkan kluster 0 dan 1 tidak berbeda signifikan. Hasil ini, bersama median HAZ yang lebih rendah, mendukung karakterisasi kluster 2 sebagai kelompok dengan stunting lebih berat. Seluruh pasangan kluster berbeda signifikan pada WAZ. Sementara itu, pada WHZ dan usia, kluster 0 berbeda signifikan dari kluster 1 dan 2, tetapi kluster 1 dan 2 tidak berbeda signifikan. Dengan demikian, WAZ membedakan seluruh kluster, HAZ terutama membedakan kluster 2, sedangkan WHZ dan usia terutama membedakan kluster 0.

E. Evaluasi Kualitas Kluster

Untuk mengevaluasi kualitas kluster yang dihasilkan oleh pendekatan PCA–DBSCAN, dilakukan perbandingan dengan dua algoritma klusterisasi pembanding, yaitu K-Medoids dan Fuzzy C-Means (FCM). Evaluasi dilakukan menggunakan tiga metrik validasi internal, yaitu *Silhouette Score*, *Davies–Bouldin Index* (DBI), dan *Dunn Index*. Perbandingan hasil evaluasi untuk masing-masing algoritma ditunjukkan pada Gambar 6.



Gambar 6. Grafik perbandingan metrik validasi internal antarmetode klusterisasi

Berdasarkan Gambar 6, pendekatan PCA–DBSCAN memperoleh nilai *Silhouette Score* tertinggi sebesar 0.277, yang menunjukkan bahwa data dalam setiap kluster memiliki kedekatan yang lebih baik serta pemisahan antarkluster yang lebih jelas dibandingkan K-Medoids (0.261) dan Fuzzy C-Means (0.260). Selain itu, nilai *Davies–Bouldin Index* (DBI) terendah sebesar 1.062 menunjukkan bahwa data pada masing-masing kluster memiliki tingkat kemiripan yang lebih baik dibandingkan K-Medoids (1.299) dan FCM (1.252). Temuan tersebut didukung oleh nilai *Dunn Index* tertinggi sebesar 0.0765, yang menunjukkan bahwa pendekatan PCA–DBSCAN menghasilkan pemisahan antarkluster yang lebih baik dibandingkan K-Medoids (0.0227) dan FCM (0.0197).

Meskipun PCA–DBSCAN memperoleh nilai validasi internal relatif lebih baik dibandingkan K-Medoids dan Fuzzy C-Means, distribusi kluster akhir menunjukkan ketidakseimbangan yang cukup besar, yaitu Kluster 0 sebanyak 357 sampel (87,93%), Kluster 1 sebanyak 24 sampel (5,91%), dan Kluster 2 sebanyak 25 sampel (6,16%). Ketidakseimbangan tersebut tidak dengan sendirinya menunjukkan kegagalan klusterisasi karena DBSCAN membentuk kluster berdasarkan struktur kepadatan data tanpa menetapkan ukuran kluster sebelumnya. Dengan demikian, perbedaan ukuran kluster dapat mencerminkan satu pola antropometri yang dominan dan dua pola yang relatif jarang pada dataset penelitian. Namun, kondisi ini perlu dipertimbangkan dalam interpretasi metrik validasi internal. *Silhouette Score* keseluruhan merupakan rata-rata nilai seluruh observasi sehingga lebih banyak dipengaruhi oleh Kluster 0 yang mendominasi sampel dan belum tentu merepresentasikan kualitas pemisahan Kluster 1 dan Kluster 2 secara setara. *Davies–Bouldin Index* mengevaluasi kompaksi dan separasi pada tingkat kluster, sedangkan *Dunn Index* bergantung pada jarak minimum antarkluster dan penyebaran maksimum intrakluster, sehingga keduanya juga perlu ditafsirkan secara hati-hati pada distribusi kluster yang sangat tidak seimbang. Selain itu, ukuran Kluster 1 dan Kluster 2 yang relatif kecil membatasi kekuatan generalisasi karakteristik kedua subkelompok tersebut. Oleh sebab itu, kedua kluster tersebut diinterpretasikan sebagai subkelompok antropometri yang ditemukan pada dataset penelitian ini dan bukan sebagai estimasi prevalensi subkelompok stunting pada populasi yang lebih luas. Perbandingan metode selanjutnya tidak hanya mempertimbangkan metrik validasi internal, tetapi juga interpretabilitas profil antropometri setiap kluster berdasarkan median HAZ, WAZ, WHZ, dan usia saat pengukuran sebagaimana disajikan pada Tabel 5.

Tabel 5. Perbandingan Profil Antropometri Kluster Berdasarkan Median

Metode	Kluster	n	HAZ	WAZ	WHZ	Usia (Bulan)
PCA–DBSCAN	0	357	–2.53	–2.12	–0.90	32.27
PCA–DBSCAN	1	24	–2.60	–0.09	1.83	18.92
PCA–DBSCAN	2	25	–3.44	–1.61	0.91	8.20
K-Medoids	0	101	–2.78	–2.97	–1.81	39.70
K-Medoids	1	227	–2.51	–1.97	–0.57	28.43
K-Medoids	2	78	–2.54	–0.87	1.03	20.40
FCM	0	143	–2.53	–2.34	–1.25	41.80
FCM	1	113	–2.51	–1.11	0.72	26.03
FCM	2	150	–2.74	–2.23	–0.99	21.05

Catatan: Nilai HAZ, WAZ, WHZ, dan usia disajikan dalam bentuk median. Kluster FCM ditentukan berdasarkan nilai keanggotaan tertinggi setiap sampel.

Perbandingan profil klinis menunjukkan bahwa K-Medoids dan FCM menghasilkan kluster yang seluruhnya memiliki median HAZ pada kategori stunting sedang. Perbedaan antarkluster pada kedua metode tersebut terutama terlihat pada WAZ, WHZ, dan usia. K-Medoids juga menghasilkan satu kluster dengan risiko gizi lebih, yaitu kluster 2 dengan median WHZ 1.03, sedangkan seluruh kluster FCM memiliki median WHZ pada kategori gizi baik. Sementara itu, PCA–DBSCAN menghasilkan kluster 1 dengan stunting sedang dan risiko gizi lebih serta kluster 2 dengan stunting berat. Secara klinis, PCA–DBSCAN memiliki kelebihan dalam mengidentifikasi kluster dengan median HAZ pada kategori stunting berat, sedangkan K-Medoids dan FCM menghasilkan distribusi anggota yang lebih seimbang tetapi tidak memisahkan profil tersebut sebagai kluster tersendiri.

IV. Kesimpulan dan saran

Penelitian ini menunjukkan bahwa integrasi *Principal Component Analysis* (PCA) dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) mampu membentuk profil keparahan stunting pada balita di Puskesmas Turikale. PCA berhasil mentransformasikan empat variabel, yaitu HAZ, WAZ, WHZ, dan usia saat pengukuran, menjadi tiga komponen utama yang tidak berkorelasi dan mampu mempertahankan 99,35% variasi data. Penerapan DBSCAN pada ruang PCA menghasilkan tiga profil akhir balita stunting, yaitu Kluster 0 yang merepresentasikan stunting sedang disertai *underweight* pada usia relatif lebih tua ($n = 357$), memerlukan pemantauan pertumbuhan berkala. Balita dengan *underweight* terkonfirmasi dapat memperoleh PMT lokal berprotein hewani dan konseling pemberian makan, disertai penilaian penyakit penyerta seperti tuberkulosis atau kecacingan berdasarkan indikasi klinis. Kluster 1 yang merepresentasikan stunting sedang disertai risiko gizi lebih pada usia lebih muda ($n = 24$), memerlukan edukasi gizi seimbang dan perbaikan kualitas asupan, pengurangan sumber gula dan lemak tidak sehat, dengan memprioritaskan protein hewani murni dan mikronutrien untuk memacu tinggi badan, serta diimbangi peningkatan aktivitas fisik sesuai tahap perkembangan anak dan pemantauan berat badan menurut tinggi badan secara berkala. Kluster 2 yang merepresentasikan stunting berat pada usia dini ($n = 25$), memerlukan rujukan untuk penilaian klinis lebih lanjut sesuai alur rujukan balita kondisi penyerta, disertai optimalisasi ASI berkelanjutan dan inisiasi MP-ASI yang tepat waktu, dan pemantauan pertumbuhan lebih ketat mengingat tingkat keparahannya. Karena penelitian hanya menggunakan data antropometri, faktor-faktor tersebut tidak dapat ditetapkan sebagai penyebab berdasarkan hasil klusterisasi. Hasil klusterisasi dapat digunakan oleh Puskesmas Turikale sebagai daftar prioritas pendamping dalam pemantauan balita melalui Posyandu. Namun, keanggotaan kluster bukan merupakan diagnosis atau dasar pemberian intervensi secara otomatis. Status gizi dan tindak lanjut setiap balita tetap harus dikonfirmasi berdasarkan klasifikasi WHO, pedoman Kementerian Kesehatan, dan penilaian klinis tenaga kesehatan. Perbandingan hasil validasi internal menunjukkan bahwa PCA–DBSCAN memberikan nilai validasi internal relatif lebih baik dibandingkan K-Medoids dan Fuzzy C-Means, dengan Silhouette Score tertinggi (0,277), Davies–Bouldin Index terendah (1,062), dan Dunn Index tertinggi (0,0765). Selain itu, PCA–DBSCAN mampu memisahkan profil dengan median HAZ kategori stunting berat sebagai kluster tersendiri, sedangkan K-Medoids dan Fuzzy C-Means menghasilkan distribusi anggota yang lebih seimbang tetapi tidak mengidentifikasi profil tersebut secara khusus. Penelitian ini memiliki keterbatasan karena menggunakan data cross-sectional yang hanya mencakup 406 balita dari satu Puskesmas pada Agustus 2025, terbatas pada indikator antropometri, serta menghasilkan distribusi kluster yang tidak seimbang dengan dua kluster berukuran kecil ($n = 24$ dan $n = 25$). Ukuran kedua kluster tersebut membatasi generalisasi karakteristik subkelompok yang ditemukan, sehingga hasilnya belum dapat dianggap merepresentasikan distribusi subkelompok stunting pada populasi yang lebih luas. Penelitian selanjutnya disarankan melibatkan beberapa Puskesmas dan memperpanjang periode pengamatan, menggunakan data longitudinal untuk menilai perubahan pertumbuhan setelah tindak lanjut, serta mengintegrasikan riwayat pertumbuhan dari e-PPGBM atau register Posyandu, riwayat penyakit dan infeksi dari rekam medis Puskesmas, praktik ASI, MP-ASI, dan pola makan dari wawancara pengasuh, serta kondisi sosial ekonomi dan sanitasi dari pendataan keluarga. Penelitian selanjutnya juga dapat membandingkan DBSCAN dengan metode berbasis kepadatan yang lebih adaptif, seperti HDBSCAN.

Ucapan Terima Kasih (Opsional)

Penulis menyampaikan terima kasih kepada Puskesmas Turikale atas izin dan akses data penelitian yang telah diberikan, serta kepada Universitas Muhammadiyah Makassar atas dukungan dan fasilitas selama proses penelitian ini. Penulis juga mengucapkan terima kasih kepada dosen pembimbing atas arahan, bimbingan, serta dukungan yang diberikan selama proses penyusunan penelitian hingga penelitian ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Y. S. Arief, F. C. Yunita, F. Efendi, F. A. K. Murti, R. O. Pradipta, and L. McKenna, “Social and Environmental Determinants of Childhood Stunting in Indonesia: National Cross-Sectional Study,” *JMIR Pediatr. Parent.*, vol. 8, no. 1, p. e68918, 2025, doi: 10.2196/68918.
- [2] World Health Organization, “The use of nutrition administrative data to generate child malnutrition estimates,” Geneva, 2025. [Online]. Available: <https://www.who.int/publications/m/item/the-use-of->

- nutrition-administrative-data-to-generate-population-level-child-malnutrition-estimates-technical-note
- [3] S. Supadmi *et al.*, “Factor related to stunting of children under two years with working mothers in Indonesia,” *Clin. Epidemiol. Glob. Heal.*, vol. 26, no. May 2023, pp. 1–6, 2024, doi: 10.1016/j.cegh.2024.101538.
 - [4] T. T. Tamir, M. A. Techane, M. T. Dessie, and K. A. Atalell, “Applied nutritional investigation spatial variation and determinants of stunting among children aged less than 5 y in Ethiopia: A spatial and multilevel analysis of Ethiopian Demographic and Health Survey 2019,” *Nutrition*, vol. 103–104, p. 111786, 2022, doi: 10.1016/j.nut.2022.111786.
 - [5] U. H. Jalloh *et al.*, “A mixed-methods study of the drivers of stunting reduction among children under-5 in Sierra Leone, 2005–2017,” *Am. J. Clin. Nutr.*, vol. 121, no. February, pp. S106–S112, 2025, doi: 10.1016/j.ajcnut.2025.02.015.
 - [6] D. Sudigyo, A. A. Hidayat, R. Nirwantono, R. Rahutomo, J. P. Trinugroho, and B. Pardamean, “Literature study of stunting supplementation in Indonesian utilizing text mining approach,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 722–729, 2022, doi: 10.1016/j.procs.2022.12.189.
 - [7] R. E. Caraka *et al.*, “Bayesian Network analysis of spatial disparities in child nutrition status and health determinants across Indonesia,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 144, no. November, pp. 1–19, 2025, doi: 10.1016/j.jag.2025.104936.
 - [8] Pemerintah Kabupaten Maros, “Fokus Entaskan Stunting di Kabupaten Maros, Bupati Maros: Prevalensi Stunting Menurun 12,3 Persen,” Kabupaten Maros. Accessed: Aug. 15, 2026. [Online]. Available: <https://maroskab.go.id/2025/08/fokus-entaskan-stunting-di-kabupaten-maros-bupati-maros-prevalensi-stunting-menurun-123-persen/>
 - [9] Christianto, “Maros Catat Penurunan Stunting Drastis, Terbesar di Sulsel,” Radio Republik Indonesia. Accessed: Aug. 16, 2026. [Online]. Available: <https://rri.co.id/makassar/regional/1762499/maros-catat-penurunan-stunting-drastis-terbesar-di-sulsel>
 - [10] T. A. E. Permatasari and Y. Chadirin, “Assessment of undernutrition using the composite index of anthropometric failure (CIAF) and its determinants: A cross-sectional study in the rural area of the Bogor District in Indonesia,” *BMC Nutr.*, vol. 8, p. 133, 2022, doi: 10.1186/s40795-022-00627-3.
 - [11] L. A. Jokhu and A. Syaury, “Determinants of concurrent wasting and stunting among children 6 to 23 mo in Indonesia,” *Nutrition*, vol. 122, p. 112390, 2024, doi: 10.1016/j.nut.2024.112390.
 - [12] B. A. Paramashanti, Y. Paratmanitya, I. I. Kusumaningtyas, T. M. Khasana, A. Yulistiyowati, and T. Siswati, “Minimum dietary diversity and the concurrence of stunting and overweight among infants and young children in Yogyakarta, Indonesia,” *Nutr. Food Sci.*, vol. 54, no. 1, pp. 120–130, 2024, doi: 10.1108/NFS-02-2023-0042.
 - [13] D. B. Rahut, R. Mishra, and S. Bera, “Geospatial and environmental determinants of stunting, wasting, and underweight: Empirical evidence from rural South and Southeast Asia,” *Nutrition*, vol. 120, p. 112346, 2024, doi: 10.1016/j.nut.2023.112346.
 - [14] N. Umwali, C. N. Kunyanga, and D. W. M. Kaindi, “Determinants of stunting in children aged between 6–23 months in Musanze region, Rwanda,” *Front. Nutr.*, vol. 9, pp. 1–12, 2022, doi: 10.3389/fnut.2022.1044350.
 - [15] W. R. Mgonezulu, P. Thangata, B. Mkandawire, and N. Amoah, “Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting,” *Hum. Nutr. Metab.*, vol. 42, no. August, p. 200340, 2025, doi: 10.1016/j.hnm.2025.200340.
 - [16] B. Rawat, H. Pant, and A. Bist, “Clustering Medical Conditions in Patient Records Using Unsupervised Learning Techniques: A Comparative Study,” *Biomed. Pharmacol. J.*, vol. 18, no. 3, pp. 2050–2060, 2025, doi: 10.13005/bpj/3236.
 - [17] G. S. Mohamed Khamis, N. S. Alqahtani, S. Munadi Alanazi, M. M. Alruwaili, M. S. Alenazi, and M. A. Alrawaili, “Using Fuzzy C-Means clustering and PCA in public health: A machine learning approach to combat CVD and obesity,” *Informatics Med. Unlocked*, vol. 57, no. August 2024, pp. 1–10, 2025, doi: 10.1016/j.imu.2025.101666.
 - [18] N. A. Rizki, A. Asyri, M. Maulidah, and C. F. Dewi, “Spatial Clustering of Urban Villages on Stunting Babies Data in Samarinda Using the DBSCAN Model,” *Media Kesehat. Masy. Indones. Vol.*, vol. 21, no. 1, pp. 24–32, 2025, doi: 10.30597/mkmi.v21i1.41250.
 - [19] A. Fahim, “Adaptive Density-Based Spatial Clustering of Applications with Noise (ADBSCAN) for Clusters of Different Densities,” *Comput. Mater. Contin.*, vol. 75, no. 2, pp. 3695–3712, 2023, doi: 10.32604/cmc.2023.036820.
 - [20] R. B. Fanda *et al.*, “The availability of essential medicines in primary health centres in Indonesia: achievements and challenges across the archipelago,” *Lancet Reg. Heal. - Southeast Asia*, vol. 22, pp.

- 1–13, 2024, doi: 10.1016/j.lansea.2023.100345.
- [21] M. Faisal *et al.*, “Assessing AI-Driven Personalization in Smart Cities Using Hybrid Machine Learning and MCDM Approach,” *HighTech Innov. J.*, vol. 6, no. 3, pp. 770–792, Sep. 2025, doi: 10.28991/HIJ-2025-06-03-03.
- [22] S. D. Darcho, F. S. Bayisa, and T. D. Nimani, “Prevalence of appropriate feeding practices and nutritional status among infant and young children aged 0–23 Months attending public hospitals in Harar town, eastern Ethiopia, 2024: An institutional-based cross-sectional study,” *Clin. Epidemiol. Glob. Heal.*, vol. 32, pp. 1–6, 2025, doi: 10.1016/j.cegh.2025.101937.
- [23] S. A. Norris *et al.*, “Infant growth and body composition from birth to 24 months : are infants developing the same ?,” *Eur. J. Clin. Nutr.*, vol. 78, pp. 952–962, 2024, doi: 10.1038/s41430-023-01386-5.
- [24] M. Greenacre, P. J. F. Groenen, T. Hastie, A. Iodice D’Enza, A. Markos, and E. Tuzhilina, “Principal component analysis,” *Nat. Rev. Methods Prim.*, vol. 2, no. 1, pp. 1–21, 2022, doi: 10.1038/s43586-022-00184-w.
- [25] D. N. Aisyah *et al.*, “Using WhatsApp for Nutrition Surveillance Among Children Under 5 Years in West Java, Indonesia: Cross-Sectional Survey and Feasibility Study,” *JMIR Pediatr. Parent.*, vol. 8, no. 1, p. e58752, 2025, doi: 10.2196/58752.
- [26] S. Tong, J. Chen, and C. Xu, “A framework for evaluating the combined effects of water transfer and storage strategies on water stress alleviation,” *Ecol. Indic.*, vol. 166, pp. 1–13, 2024, doi: 10.1016/j.ecolind.2024.112542.
- [27] S. Oprea and B. Adela, “Heliyon Profiling public perception of emerging technologies : Gene editing , brain chips and exoskeletons . A data-analytics framework,” vol. 10, pp. 1–22, 2024, doi: 10.1016/j.heliyon.2024.e40268.
- [28] A. Keramatian, V. Gulisano, and M. Papatriantafilou, “IP.LSH.DBSCAN: Integrated parallel density-based clustering by locality-sensitive hashing,” *Discret. Appl. Math.*, vol. 382, pp. 183–196, 2025, doi: 10.1016/j.dam.2025.11.047.
- [29] K. Giri, T. K. Biswas, and P. Sarkar, “ECR-DBSCAN: An improved DBSCAN based on computational geometry Kinsuk,” *Mach. Learn. with Appl.*, vol. 6, no. August, p. 100148, 2021, doi: 10.1016/j.mlwa.2021.100148.
- [30] D. Y. Badawood, “StatVis : A visual analytics framework for statistical cluster validation in high dimensions,” *Array*, vol. 28, pp. 1–16, 2025, doi: 10.1016/j.array.2025.100560.
- [31] A. N. Belkacem and A. Hireche, “Exploring student engagement and learning preferences: A comparative study between virtual- and robot-based tutors,” *Meas. Sensors*, vol. 38, pp. e1–e5, 2025, doi: 10.1016/j.measen.2024.101704.
- [32] M. Monshizadeh, V. Khatri, R. Kantola, and Z. Yan, “A deep density based and self-determining clustering approach to label unknown traffic,” *J. Netw. Comput. Appl.*, vol. 207, pp. 1–18, 2022, doi: https://doi.org/10.1016/j.jnca.2022.103513.
- [33] H. Ghaderi, B. Foreman, A. Nayebi, S. Tipirneni, C. K. Reddy, and V. Subbian, “A self-supervised learning-based approach to clustering multivariate time-series data with missing values (SLAC-Time): An application to TBI phenotyping,” *J. Biomed. Inform.*, vol. 143, 2023, doi: 10.1016/j.jbi.2023.104401.
- [34] N. Pakgohar, A. Lengyel, and Z. B. Dukat, “Quantitative evaluation of internal cluster validation indices using binary data sets,” *J. Veg. Sci.*, vol. 35, p. e13310, 2024, doi: 10.1111/jvs.13310.
- [35] A. N. Brzezinska and I. Gaibei, “How the Outliers Influence the Quality of Clustering?,” *Entropy*, vol. 24, no. 7, 2022, doi: 10.3390/e24070917.
- [36] A. Rafisa, E. Sarilita, B. Delage, R. G. Munger, and P. A. Mossey, “Situational analysis of nutritional status among 1899 children presenting with cleft lip and/or palate in Indonesia,” *J. Glob. Health*, vol. 13, p. 04127, 2023, doi: 10.7189/jogh.13.04127.
- [37] B. N. Damanik and H. Hely, “Determinants of Hospital Length of Stay Among Pulmonary Tuberculosis Inpatients at A Secondary General Hospital in Indonesia: A Comparative Analysis,” *J. Med. Informatics Technol.*, vol. 4, no. 2, pp. 64–69, 2026, doi: 10.37034/medinftech.v4i2.150.
- [38] M. Thamrin *et al.*, “Machine learning for global trade analysis: a hybrid clustering approach using DBSCAN, elbow, and SOM,” *IAES Int. J. Artif. Intell.*, vol. 14, no. 4, pp. 3033–3046, 2025, doi: 10.11591/ijai.v14.i4.pp3033-3046.
- [39] M. Faisal *et al.*, “Utilizing Machine Learning-Based Decision-Making to Align Higher Education Curriculum with Industry Requirements,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 17, no. 4, pp. 1–25, 2025, doi: 10.5815/ijmecs.2025.04.01.